Ear-to-ear Capture of Facial Intrinsics

Seck, Alassane; Dutta, Abhishek; Smith, William; Tiddeman, Bernard; Dee, Hannah; Dessein, Arnaud

Publication date:
2016

Citation for published version (APA):
Seck, A., Dutta, A., Smith, W., Tiddeman, B., Dee, H., & Dessein, A. (2016). *Ear-to-ear Capture of Facial Intrinsics*. arXiv. https://arxiv.org/abs/1609.02368

Ear-to-ear Capture of Facial Intrinsics

Seck, Alassane; Dutta, Abhishek; Smith, William; Tiddeman, Bernard; Dee, Hannah; Dessein, Arnaud

DOI:
https://arxiv.org/abs/1609.02368

Publication date:
2016

Citation for published version (APA):
Seck, A., Dutta, A., Smith, W., Tiddeman, B., Dee, H., & Dessein, A. (2016). *Ear-to-ear Capture of Facial Intrinsics*. arXiv. https://doi.org/https://arxiv.org/abs/1609.02368

Ear-to-ear Capture of Facial Intrinsics

Alassane Seck, Abhishek Dutta, William A. P. Smith, *Member, IEEE,*
Bernard Tiddeman, Hannah Dee, and Arnaud Dessein

Abstract—We present a practical approach to capturing ear-to-ear face models comprising both 3D meshes and intrinsic textures (i.e. diffuse and specular albedo and normal maps). Our approach is a hybrid of geometric and photometric methods and requires no geometric calibration. Photometric measurements made in a lightstage are used to estimate view dependent high resolution normal maps. We overcome problems of fixed photometric viewpoint by capturing in multiple poses. We use a multiview reconstruction pipeline of structure-from-motion followed by multiview stereo to compute a base mesh to which the photometric views are registered. We propose a novel approach to robustly stitching the normal maps and intrinsic textures into a seamless, complete and detailed face model. The resulting models provide photorealistic renderings in any view.

Index Terms—Diffuse albedo, face capture, multiview stereo, photometric stereo, specular albedo.

1 Introduction

Measuring properties of a face that are truly intrinsic (i.e. unrelated to environmental or imaging parameters) is a longstanding goal with applications in a wide range of fields. In graphics, it allows face images to be synthesised in arbitrary illumination conditions [1], simulating the properties of any camera. In statistical modelling, it allows the variability in a population of faces to be studied independently of imaging conditions [2]. In computer vision, it allows appearance in an image to be predicted for pose and illumination invariant recognition or classification [3], [4]. In psychology, it allows the relative importance of shape and intrinsic texture to the neural representation of faces to be studied [5]. Despite these broad and compelling applications, and over a decade of research attention, there remains no satisfactory method for capturing intrinsic properties of a whole face (from ear-to-ear).

By "intrinsic properties" we refer specifically to the shape and reflectance properties of a face that give rise to a particular face appearance when illuminated and imaged. Shape is usually represented by a 3D mesh and reflectance properties by 2D parameter maps representing the spatial distribution of reflectance parameters in texture space. In turn, reflectance parameters are determined by the spatial distribution of biophysical parameters such as skin pigmentation and facial hair over the face surface.

A face is only partially visible from a single viewpoint. In any single view, parts of the face are either occluded or so foreshortened that their projected resolution is too low to provide useful information. However, for many applications a full face model is required. For example, it has been shown that the ears are an important feature for 3D face modelling [6]. Likewise, cropped face models introduce artificial boundaries that make it difficult to use the model as part of a character animation or may disrupt neural processes when used as psychological stimuli.

In this paper we present an approach that enables the intrinsic shape and reflectance properties of a face to be captured over the whole face.

- *A. Seck, B. Tiddeman and H. Dee are with the Department of Computer Science, Aberystwyth University, UK. Email: {als31,bpt,hmd1}@aber.ac.uk*
- *W. Smith and A. Dessein are with the Department of Computer Science, University of York, UK. Email: {william.smith,arnaud.dessein}@york.ac.uk*
- *A. Dutta is with the University of Twente, Netherlands. Email: a.dutta@utwente.nl*

1.1 Related Work

Existing methods for face shape capture fall broadly into two categories: photometric and geometric. Photometric methods use the intensity of reflected light to infer the orientation and material properties of the face surface. Geometric methods use feature point positions observed from multiple viewpoints to infer the depth of the face surface. The advantage of photometric methods is that they are dense (measurements are made at every pixel and resolution is limited only by the resolution of the camera). Moreover, photometric analysis allows estimation of additional reflectance properties such as diffuse and specular albedo [1], surface roughness [7] and index of refraction [8]. This information is essential for rendering or relighting the captured face models.

However, surface orientation is only a 2.5D shape representation and the estimated normal field must be integrated in order to recover surface depth [9] or used to refine a 3D mesh captured using other cues [10]. Also, photometric methods are usually much more demanding in data capture terms and also in their requirement for controlled conditions.

Geometric methods on the other hand allow instantaneous capture of face geometry but do not allow

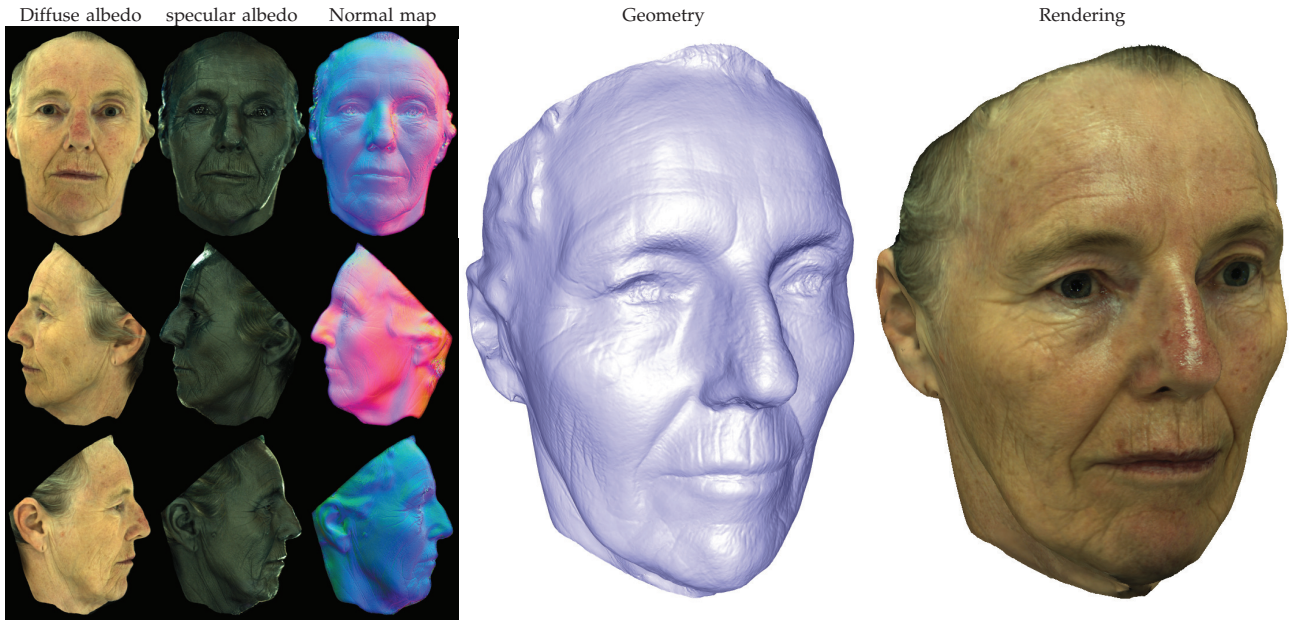


Fig. 1. Ear-to-ear face capture

estimation of reflectance properties, providing only a fixed texture map. Below, we review existing methods for photometric and geometric face capture as well as hybrid methods and models of skin reflectance that are used for face analysis and rendering.

Photometric methods. Photometric shape estimation has a long history in computer vision [11]. Recently, there has been a resurgent interest in applying photometric methods to the problem of face capture [1], [7], [8], [12]. In particular, the spherical gradient-based photometric stereo method of Ma et al. allowed finescale facial features such as skin pores and wrinkles to be recovered with enhanced accuracy and robustness in comparison to traditional point source methods. This was extended to realtime performance capture by Wilson et al. who showed how a certain sequence of illumination conditions allowed

A lightstage uses polarisation to separate specular and diffuse reflectance. This is most easily achieved by placing a linear polarising filter in front of each point source. The filter is oriented such that, once reflected specularly from the face, the plane of polarisation is the same for all sources. Unfortunately, the required orientation is viewpoint dependent: a given calibration only separates diffuse and specular reflection for two (antipodal) viewing directions. Hence, there is no straightforward way to perform multiview photometric analysis using a lightstage.

Going further, even using a lightstage, albedo maps measured as tristimulus (RGB) images are tied to the spectral sensitivities of the camera and spectral power distribution of the light source. This means that a captured face cannot be composited into a new scene without careful colour transformation. Moreover, only certain colours can plausibly arise from the pigments and layered structure of skin [13], a constraint not enforced. Using the approach of Ma et al. the albedo maps are further corrupted by "ambient occlusion" and inter-reflection effects meaning shape and material properties are not fully separated.

Geometric methods. Multiview shape estimation methods such as binocular stereo, structure-from-motion and multiview stereo have been applied quite successfully to the problem of face shape estimation. The key problem in this context is establishing correspondence between views over apparently-featureless regions of the face such as the cheeks and forehead. One solution to this problem is to either paint [14] or project [] a pattern onto the face that provides matchable features. An alternative passive approach is to use very high resolution images in which fine scale features such as freckles, wrinkles and skin pores are resolved. These provide ideal features for robust matching. Finally, methods such as space carving do not rely on feature matching and can hence be applied to faces [] even when small features are not visible.

The state-of-the-art approach in geometric face capture is due to Beeler et al. [15]. Since their method is reliant on a very accurate geometric calibration, they propose a novel calibration process based on a spherical calibration target. Shape estimation then proceeds in two steps. First a base mesh is obtained using a multiview stereo approach. Next, detail is embossed onto the mesh using a shading-based heuristic. Whilst the resulting meshes contain convincing detail, the fine scale shape detail is not accurate since it is hallucinated from a texture cue rather than satisfying any meaningful geometric or photometric constraint.

In general, any multiview method that relies on accurate intrinsic and extrinsic calibration is highly restrictive. The camera focus must be fixed between calibration and face capture. This is particularly prob-

lematic if such a setup is to be integrated with a photometric system. For example, in a lightstage the amount of light received by the camera is limited by the polarising filters on illuminants and camera. This usually means that a relatively large aperture is used, leading to a reduced depth of field. In such a case, focus is very sensitive and it is unlikely that a single focus would suffice for both calibration and capture.

Of course, purely geometric methods can only recover one intrinsic property of faces: shape. Texture maps are nothing other than a photograph of the face under a particular set of environmental conditions. Hence, the texture obtained using multiview stereo is useless for relighting. Worse, since appearance is view-dependent (the position of specularities changes with viewing direction), no one single appearance can explain the set of multiview images.

Hybrid methods There have been a number of attempts to combine photometric and geometric methods for face or object capture. This is largely motivated by the fact that their advantages complement the weaknesses of the other approach.

Nehab et al. [10] proposed an efficient approach for combining estimated surface normals and surfaces (in the form of a depth map or mesh). Their approach is particularly applicable to surface normals estimated using photometric methods which are likely to contain low frequency bias. This low frequency bias is removed by the base mesh which in return is refined by the accurate high frequency detail present in the photometric surface normal. Their approach is based on a linear approximation to the underlying objective of minimising angular error between target normals and those of the final surface.

Wu et al. [16] propose an approach that combines multi-illumination MVS and uncalibrated photometric stereo methods. They recover depth maps from multi-view and multi-illumination video sequences, then merge these to a watertight mesh using multi-illumination photo-consistency constraints. The recovered mesh is further refined with photometric surface normals measured under uncalibrated (unknown) illumination conditions. The proposed uncalibrated photometric stereo technique consist in an iterative estimation of both the surface normals and the illumination conditions simultaneously. The authors first initialize the surface normals with the ones from the mesh produced by MVS, and then minimize the shading errors assuming a spherical harmonic representation of the lighting. However, while this approach has the advantage of not requiring controlled illumination, it considers only Lambertian surfaces and tends to fail in presence of cast shadows or strong inter-reflections.

The closest previous work to what we propose in this paper is due to Ghosh et al. [17]. They capture multiple views photometric data simultaneously in a lightstage. In order to overcome the view dependency of the polariser orientation calibration, they make an empirical observation. Namely, using two illumination fields with locally orthogonal patterns of polarisation (i.e. filters aligned to lines of latitude or longitude) allows approximate specular/diffuse separation from any view close to the equator. Unfortunately, this approximation means that their approach does not recover truly intrinsic properties and therefore does not fulfil our goals. Specular and diffuse reflectance is not fully separated meaning that both normal maps and specular/diffuse albedo maps are corrupted. In addition, they propose an empirical process for "Fresnel compensation" of the specular albedo maps. As we discuss in Section X, this is based on an assumption of constant specular reflectance properties and neglects rough surface effects. This further reduces the utility of the specular albedo as an intrinsic quantity.

Park et al. [18] propose avoiding the multiple views photometric merging problem by computing the surface normals directly in the 2D parameter domain of the mesh. They start with using SFM techniques to compute a base mesh which they then parameterize to a piecewise planar space. The parameterization is used to wrap the multiple view/illumination images onto the base mesh; and uncalibrated multi-view photometric stereo [19] is used to estimate the surface normal directly in the 2D parameter domain. Finally the base geometry is refined using displacement maps calculated in the 2D parameter domain form the estimated surface normals. The authors use linear reflectance model meaning that the proposed method will work only for Lambertian surfaces.

1.2 Contributions

We begin by proposing a novel capture pipeline that uses uncalibrated multiview stereo to register photometric views to a base mesh. Our proposed approach enables photometric measurements to be made over the whole face. We then revisit the spherical gradient photometric stereo approach of Ma et al. [1] and Wilson et al. [?]. We also provide a photometric alignment method that avoids the iterative approach of Wilson et al. . Finally, we present a robust approach to stitching photometric information across the multiple views, accounting for the view-dependent low frequency bias present in the estimated normals. This includes a novel, unified approach to stitching both textures and shape using screened Poisson equations.

Our approach requires no geometric calibration, no inverse rendering, only consumer hardware and is fast (the whole capture process comprises three sequences lasting around 2 seconds each). Yet, the quality of the estimated shape is comparable to state-of-the-art methods that require much more careful calibration, e.g. [15]. Moreover, our method also estimates diffuse and specular reflectance maps meaning our models are relightable.

2 Pipeline

The polarisation properties of light have been widely used as a cue to study surface shape and reflectance properties. One of the best known effects is that specular reflectance from a dielectric material preserves the plane of polarisation of linearly polarised incident light. The linear polarisation is lost by diffuse reflectance. This means that specular and diffuse reflectance can be separated by illuminating an object with a linearly polarised light source and then placing another linear polariser in front of the camera. If the camera's filter is oriented perpendicular to the light source's filter, then specular reflectance is entirely eliminated. If they are parallel, both specular and diffuse reflectance is observed. The specular reflectance is obtained by taking the different between the two images.

In a lightstage, a face is illuminated by light sources distributed over a geodesic sphere. Each light source must have the orientation of its polarising filter (and hence its plane of polarisation) set such that a specular reflection towards the viewer leads to all reflections having the same plane of polarisation. This means that a single linear polarising filter on the camera can separate specular and diffuse reflectance from all light sources simultaneously. However, while this calibration is relatively straightforward it is, unfortunately, view dependent. A given calibration only separates diffuse and specular reflection for two (antipodal) viewing directions.

We overcome this problem by capturing a face multiple times in different poses relative to the calibrated viewpoint, e.g. frontal and two profile views. Together, these three photometric views provide full ear-to-ear coverage of the face. We augment the photometric camera with additional cameras providing multiview, single-shot images captured in sync with a reference frame of the photometric sequence (the diffuse constant image). We position these additional cameras to provide overlapping coverage of the face. Since we do not rely on a fixed calibration, their exact positioning is unimportant and we allow the cameras to autofocus between captures. In our setup, we use 7 such cameras giving a total of 8 simultaneous views. Since we repeat the capture three times, we have 24 effective views.

To merge these views and to provide a rough base mesh, we perform a multiview reconstruction using the photometric views augmented by additional cameras providing multiview, single-shot images captured in sync with the constant illumination condition. Solving this uncalibrated multiview reconstruction problem provides

Note that since the three photometric views are not acquired simultaneously, there is likely to be non-rigid deformation of the face between these views. For this reason, in Section 5 we propose a robust algorithm for stitching the views without blurring potentially misaligned features.

1) **Multiview reconstruction**: We commence by applying structure-from-motion followed by multiview stereo to
2) **View merging**

The photometric views will capture spherical gradient sequences [1], [12], however we will use opto-electrical polarising filters to allow diffuse/specular separation without the need for mechanical filter rotation. These views will be augmented by additional cameras providing multiview, single-shot images captured in sync with the constant illumination condition.

3 Spherical Gradient Photometric Stereo

Spherical Gradient Photometric Stereo has been well studied in [1], [12]. Their idea amounts to something very simple: estimate the first moment (centre of mass) of the reflectance lobe at a point by illuminating that point with a linear spherical gradient. For a Lambertian surface, this direction coincides with the surface normal and, for a specular surface, with the reflection direction (from which the surface normal can be calculated).

3.1 Background

3.1.1 The Lambertian case

Let r_x^d, r_y^d, r_z^d and r_c^d respectively the measured Lambertian radiances under the $X-$gradient, $Y-$gradient, $Z-$gradient and constant illuminations, [1] established the relation between the surface normal $n^d = (n_x^d, n_y^d, n_z^d)$ and the measured Lambertian radiance as follow:

$$\begin{aligned} n_x^d &= \frac{1}{N_d}\left(\frac{r_x^d}{r_c^d} - \frac{1}{2}\right) \\ n_y^d &= \frac{1}{N_d}\left(\frac{r_y^d}{r_c^d} - \frac{1}{2}\right) \\ n_z^d &= \frac{1}{N_d}\left(\frac{r_z^d}{r_c^d} - \frac{1}{2}\right), \end{aligned} \tag{1}$$

where, N_d is a normalizing constant given by

$$N_d = \sqrt{\left(\frac{r_x^d}{r_c^d} - \frac{1}{2}\right)^2 + \left(\frac{r_y^d}{r_c^d} - \frac{1}{2}\right)^2 + \left(\frac{r_z^d}{r_c^d} - \frac{1}{2}\right)^2} \tag{2}$$

3.1.2 The specular case

In the specular case, [1] show that it is easier to estimate, from the measured specular radiances, the specular reflection vector than the surface normal directly. If r_x^s, r_y^s, r_z^s and r_c^s denote respectively the measured specular radiances under $X-$gradient, $Y-$gradient, $Z-$gradient and constant illuminations, the specular reflection vector $u = (u_x, u_y, u_z)$ is given by:

Fig. 2. The set of 63 images used in our pipeline. We obtain 24 different viewpoints of the face (boxed blue images) by capturing 8 different views with the subject in a frontal (rows 1 and 2), left profile (rows 3 and 4) and right profile (rows 5 and 6). Rows 1, 3 and 5 show the sequence of images captured for the photometric viewpoint. Rows 2, 4 and 6 show the multiview images. Images within a blue box are captured simultaneously. The remainder are captured in sequence from left to right.

$$\begin{aligned} u_x &= \frac{1}{N_s}\left(r_x^s - \frac{1}{2}r_c^s\right) \\ u_y &= \frac{1}{N_s}\left(r_y^s - \frac{1}{2}r_c^s\right) \\ u_z &= \frac{1}{N_s}\left(r_z^s - \frac{1}{2}r_c^s\right) \end{aligned} \tag{3}$$

where, N_s is a normalizing constant given by

$$N_s = \sqrt{\left(r_x^s - \frac{1}{2}r_c^s\right)^2 + \left(r_y^s - \frac{1}{2}r_c^s\right)^2 + \left(r_z^s - \frac{1}{2}r_c^s\right)^2} \tag{4}$$

As the surface normal corresponds to the direction half-way between the view vector v ($v = (0,0,1)$ in our case) and its specular reflection u, it can be obtained by:

$$\vec{n} = \frac{1}{\bar{N}}(u+v) \tag{5}$$

Where $\bar{N}$ is a normalizing vector.

3.1.3 Complement Gradient Illumination

[12] proposed a more optimal way of calculating the surface normals from Spherical Gradient Illumination. The authors exploit the the gradient images complements obtained under the complementary lighting conditions. Thus, in addition to the four gradient images r_x,r_y, r_z and r_c proposed in [1], they captured three others $r_{\bar{x}}$,$r_{\bar{y}}$ and $r_{\bar{z}}$ such that:

$$r_x + \bar{r}_x = r_y + \bar{r}_y = r_z + \bar{r}_z = r_c \tag{6}$$

From 6, 1 and a re-normalization, they obtain:

$$n = \frac{[r_x - \bar{r}_x, r_y - \bar{r}_y, r_z - \bar{r}_z]^T}{||\,[r_x - \bar{r}_x, r_y - \bar{r}_y, r_z - \bar{r}_z]^T\,||} \tag{7}$$

This method is proven to improve the quality of the calculated normals and is more robust to image quality than the method in [1]. This is explained by the fact that the dark regions in one gradient image are likely to be well lit in the complement image.

4 Photometric Alignment

Since spherical gradient photometric stereo requires a set of images to be captured in series, the images

within a sequence may not be in perfect alignment due to subject motion. In the context of estimating fine scale shape, these small misalignments lead to a blurring of detail. Since inter-frame motion is likely to be very small (perhaps sub-pixel) and visibility is unlikely to change between views, the obvious solution is to use optical flow to align each image to a reference frame. However, since the illumination changes in each frame the usual brightness constancy constraint does not apply (we expect the brightness of a given point on the face to vary dramatically as illumination changes).

Wilson et al. [12] overcame this problem by exploiting a property of the complement images. Assuming no motion, the sum of a gradient image and its complement are equal to the constant image. Hence, an alternative brightness constancy constraint can be written down. For example, for the x-gradient images:

$$C(x,y) = X(x+\Delta x_1, y+\Delta y_1) + \bar{X}(x+\Delta x_2, y+\Delta y_2). \quad (8)$$

This involves solving for the optical flow vectors for both gradient and complement images in one go. Wilson et al. [12] propose an iterative approach to this problem where they initially compute the flow from X to $C-\bar{X}$ followed by the flow from $\bar{X}$ to $C-w(X)$, where w is the warp computed at the previous step. It is propose that iterating these two steps converges towards the correct flow for both images.

A weakness of their approach is that $C-\bar{X}$ is not necessarily a good target for warping onto. Since they are not aligned, taking their difference leads to a blurring of features to which X is unlikely to be satisfactorily warped. In an extreme case, it can be shown that this method can fail completely.

We propose an alternative that is both more efficient and more robust. We note that changing the spherical illumination pattern affects only intensity and not colour. Thus we use color space transformations to extract intensity-free information from images with different illumination condition. Hue-Saturation-Value(HSV) and normalized-RGB color spaces are known to be efficient ways of separating intrinsic color from shading related-intensity [20]. For an image I, we combine the Hue component of the HSV space with normalized-RGB to produce an illumination-independent image I_{color}:

$$I_{color} = \frac{1}{2}\left\{ hue(I) + \frac{I}{\sqrt{I_R^2 + I_G^2 + I_B^2}} \right\} \quad (9)$$

Figure 3 shows two images in different spherical gradient lighting patterns (X-gradient and Constant) and the corresponding illumination-independent images.

However, while allowing good alignments on global shapes, the color transformation tends to smooth out fine details which can lead to local misalignments. These are more significant as the motion

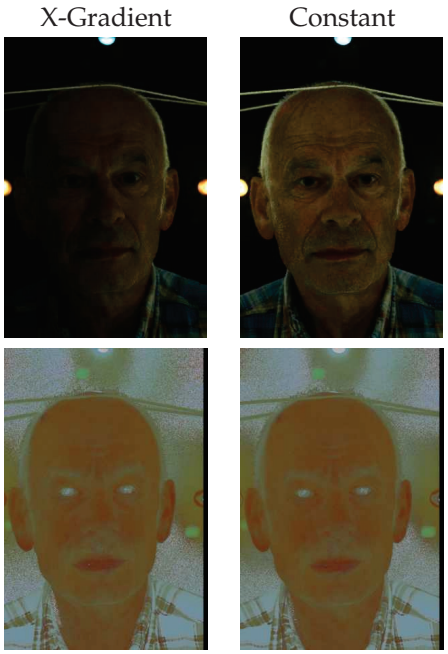


Fig. 3. Illumination-independent images for Photometric Alignment

is not rigid. We correct this by employing our method to initialize Wilson's method: we use the flow between C_{color} and $\bar{X}_{color}$ to align C and $\bar{X}$ before computing $C-\bar{X}$. In practice, our experiments show that only one iteration after initialization is enough to get very good alignments. Figure 4 compares normal-maps obtained when photometric images are aligned with 3 iterations of Wilson's method (a) and only 1 iteration when initialized with our method (b).

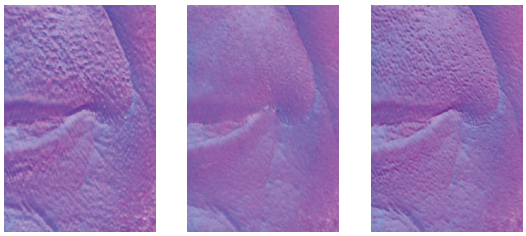


Fig. 4. Normal-maps obtained with different alignment strategies. (a) Wilson's method after 3 iterations. (b) Our method after 0 iteration (only initialization). (c) Our method after 1 iteration

5 Stitching Photometric Views

In this section we describe a method for seamlessly stitching the intrinsic textures and normal maps from each photometric view onto the base mesh obtained with multiview stereo. This is a non-trivial problem.

The constraints of linear polarisation necessitate that the three photometric image sets are taken at different times (the subjects rotate themselves to allow capture of frontal and two profile views). Hence, the face is likely to have changed shape between views meaning that there is no single correct shape and

correspondence between images and mesh is imperfect. In addition, certain reflectance effects introduce a view-dependency on the intrinsic textures. For example, Fresnel gain means that specular albedo is unreliable close to occluding boundary (see Figure 1). Applying a baseline stitching algorithm (such as back-projection and averaging) to such data leads to blurring of misaligned features, visible seams between textures taken from different views and inclusion of unwanted specular effects.

To address these problems, we propose a unified approach that allows us to stitch both intrinsic textures and shape. Our approach is based on Poisson blending using non-conservative guidance fields. The guidance fields are either in the form of texture gradients or photometric surface normals. Our approach is patch-based. The idea is to blend the sampled images by least-angle selection of gradients in overlapping patches. The majority of texture stitching algorithms are vertex- or face-based strategies with additional heuristics for robustness. We expect a patch-based approach to improve robustness by allowing selection criteria to be aggregated over a patch. Also, since a patch is taken from a single view, there will be no blending artefacts within a patch while the patch overlaps provide a means to blend between textures taken from different views.

Since we rely on discrete differential operators on the mesh surface, our approach completely preserves conservative vector fields compared to extrinsic 3D finite elements in [21]. This makes our approach more natural from a theoretical perspective, even if non-conservative fields are rather formed in practice.

5.1 Poisson Blending

Blending in the gradient domain via solution of a Poisson equation was first proposed by Pérez et al. [22] for 2D images. The motivation is that second order variations in texture are the most significant perceptually whereas low frequency texture variations have a barely noticeable effect. The same argument can be made for texture and geometry on a mesh. The approach allows us to avoid visible seams where the texture or geometry from two different views are inconsistent.

Hence, the idea is to form a guidance field of texture gradients $\mathbf{v}$ selected from source images and then solve for the texture f whose gradients best match the guidance field:

$$\min_f \int\int_\Omega \|\nabla f - \mathbf{v}\|^2 dA \tag{10}$$

This minimisation problem can be solved by solving the Poisson equation:

$$\Delta f = \nabla \cdot \mathbf{v}, \tag{11}$$

where Δ is the Laplace operator and $\nabla\cdot$ is the divergence operator.

5.2 Discrete differential operators

In order to solve a Poisson equation over a triangle mesh, we need to define discrete counterparts to the Laplace and divergence operators. We consider a triangular base shape mesh $\mathcal{M}$ and assume it describes a 2D manifold S. The connectivity is given by a simplicial complex $\mathcal{K}$ whose elements are vertices $\{i\}$, edges $\{i,j\}$ or faces $\{i,j,k\}$, with indices $i,j,k \in [1\,..\,N]$, where N is the number of vertices. We write a vertex $\{i\}$ as i for simplicity.

A discrete vector field V is a piecewise constant vector function defined for each triangle T_l by a coplanar vector v_l. A discrete potential field is a piecewise linear function $\phi(s) = \sum_{i\in\mathcal{K}} \phi_i B_i(s)$ on the mesh surface, where B_i is the piecewise linear basis function valued 1 at vertex i and 0 at other vertices, and ϕ_i specifies the value of ϕ at vertex i. The discrete gradient of ϕ for triangle T_l is $\nabla\phi_l = \sum_{i\in\mathcal{K}} \phi_i \nabla B_{il}$, where ∇B_{il} is the gradient of B_i within T_l. The divergence of V at vertex i is $\operatorname{div} V(i) = \sum_{T_l\in\mathcal{K}_i} |T_l|\, \nabla B_{il}^\top\, \mathrm{v}_l$, where $\mathcal{K}_i$ is the set of triangles sharing vertex i and $|T_l|$ is the area of triangle T_l. Writing Poisson's equation $\operatorname{div}\nabla\phi = \operatorname{div} V$ in this framework leads to a linear system of equations $\mathbf{Ax} = \mathbf{y}$ for the unknown potential values $x_i = \phi_i$, where:

$$a_{ij} = \sum_{T_l\in\mathcal{K}_i} |T_l|\, \nabla B_{il}^\top \nabla B_{jl} \;, \quad y_i = \sum_{T_l\in\mathcal{K}_i} |T_l|\, \nabla B_{il}^\top\, \mathrm{v}_l \;. \tag{12}$$

This system is sparse since the sum for coefficients a_{ij} is non-null iff $\{i,j\} \in \mathcal{K}$ (it is an edge). The sum is then simply over the triangles T_l (two if not a boundary edge, one otherwise) sharing this edge.

This equation can be interpreted as seeking for a potential field ϕ whose gradient $\nabla\phi$ matches the guide vector field V. If V is conservative, i.e., it is the gradient of an existing potential field ϕ, then ϕ is the exact solution. Otherwise, a more general minimizer can still be obtained by least squares but its gradient differs from V. In addition, we regularize the minimization via screening:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x} - \mathbf{x}'\|_2^2 \;, \tag{13}$$

where $\lambda > 0$ and $\mathbf{x}'$ defines a guide potential field ϕ' as $\phi'_i = x'_i$.

5.3 Image sampling

For each view, we determine the set of visible vertices on the 3D model. Occlusions can be tested by ray casting from the viewer, although here we use an approximate but much less demanding depth criterion with a z-buffer. The image is then sampled on the mesh by back-projection of textures for visible vertices after bilinear color interpolation within the pixel grid. Additionally, the viewing angles for each face and vertex are computed as part of the process

and stored for later use. Any additional heuristics related specifically

We notice finally that any of the heuristics discussed in the introduction could be employed here to discard some vertices that are potentially corrupted by considering them as occluded.

5.4 Mesh segmentation

We achieve mesh segmentation with a classical farthest-point strategy [23], enhanced with an original patch growing scheme to form an overlapping structure. This produces a uniform segmentation compared to the region-growing scheme in [24]. Moreover, we do not require patches to be disk-homeomorphic since our methods are intrinsic to the mesh surface. Not only does it eliminate undesirable distortions inherent to flattening, but it also allows to consider various mesh topologies with arbitrary genus and number of boundary components. Lastly, instead of using extrinsic 3D ball radii as thresholds for patch dilation, overlaps are obtained by growing patches intrinsically within neighbors via geodesic projections.

We first select vertices iteratively by adding a new sample one at a time. Our mesh is equipped with a geodesic distance map D. Denoting by $D_l(i)$ the geodesic distance map to the first l selected samples, we select sample $i^\star_{l+1}$ as the vertex that maximizes $D_l(i)$. The distance map $D_{l+1}(i)$ can simply be updated as the minimum between $D_l(i)$ and $D(i, i^\star_{l+1})$. We continue this process until a desired number M of vertices have been sampled.

Patches $\mathcal{P}_1, \ldots, \mathcal{P}_M$ are then obtained via the geodesic Voronoi tessellation based on the samples. The segmentation thus defines a dual graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = [1\,..\,M]$, and $(m, n) \in \mathcal{E}$ if $\mathcal{P}_m$ and $\mathcal{P}_n$ are neighbors, i.e., are connected by an edge $\{i, j\} \in \mathcal{K}$. To grow a patch $\mathcal{P}_m$, we consider separately each of its neighbor patches $\mathcal{P}_n$ with $(m, n) \in \mathcal{E}$, and define thresholds d_{mn} as follows:

$$d_{mn} = \sigma \times D(i^\star_m, i^\star_n) \ , \tag{14}$$

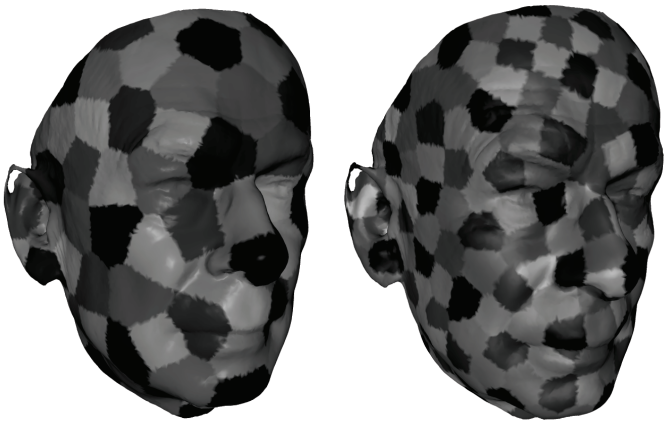
where $\sigma \geq 0$ is set by the user and can be seen as an overlap ratio or factor, and the geodesic distance D is restricted to the union $\mathcal{P}_m \cup \mathcal{P}_n$ of the reference patch and considered neighbor. The overlap $\mathcal{O}_{mn}$ of $\mathcal{P}_m$ onto $\mathcal{P}_n$ is then constructed by geodesic projections:

$$\mathcal{O}_{mn} = \left\{ i \in \mathcal{P}_n \colon \min_{j \in \mathcal{P}_m} D(i, j) \leq d_{mn} \right\} \ . \tag{15}$$

A given grown patch $\mathcal{Q}_m$ is eventually constructed by concatenation of the reference patch $\mathcal{P}_m$ with the respective overlaps:

$$\mathcal{Q}_m = \mathcal{P}_m \cup \bigcup_{n|(m,n)\in\mathcal{E}} \mathcal{O}_{mn} \ . \tag{16}$$

The time complexity of the whole process is in $O(N \log N \log M)$.



(a) (b)

Fig. 5. Mesh segmentation with different sampling vertices number (left:100; right:400)

5.5 Texture blending

We apply this to solve for texture by considering each color channel independently as a potential field ϕ. For each view v, we compute the mean viewing angle of vertices in the different patches. Unobserved vertices, due either to occlusion or missing information, are assumed to have a viewing angle of $\pi/2$. Hence, patches with unobserved data are penalized and no difference on the nature of non-observability is made. For each patch now, we select texture from the view where the patch has the smallest viewing angle. For unobserved vertices, we also select texture from subsequent sorted views. We end up with partial textures $\phi^{(v)}$ that we stitch in overlaps by Poisson blending. To build up the guide vector field V, we select local texture gradients by least angle for each triangle T_l:

$$\mathrm{v}_l = \sum_{i \in \mathcal{K}} \phi_i^{(v_l)} \nabla B_{il} \ , \tag{17}$$

where v_l is the view whose angle is minimal for triangle T_l. We also fill in unobserved faces simply by setting their gradients to zero for smoothness. Screening is done via a rough estimate ϕ' obtained by averaging textures $\phi^{(v)}$, unobserved textures being discarded from the regularization. We use a small penalty λ=10^{-6} to remove color offset indeterminacies since we did not observe dramatic color bleeding issues compared to [21]. The time complexity for building up the linear system is in $O(M + N)$. A naive complexity for least squares optimization via Cholesky decomposition is in $O(N^3)$, though efficient solvers that exploit sparsity can be used instead.

We show in figure 6 the results of the gradient stitching on the diffuse and specular textures.

5.6 Surface normal blending

From differential geometry, it is known that:

$$\lim_{|\gamma|\to 0} \frac{1}{|\gamma|} \int_{\mathbf{v}\in\gamma} (\mathbf{v}_i - \mathbf{v}) dl(\mathbf{v}) = -H(\mathbf{v}_i)\mathbf{n}_i \tag{18}$$

(a) (b)

Fig. 6. Results of our gradient stitching (a) diffuse (b) specular

Ultimately, our goal is to transfer the detail from the photometric normal maps to the mesh surface. One approach to this problem would be to start by stitching the normal maps into a seamless and complete normal map for the whole face. Then, the normals could be embossed onto the mesh using an algorithm such as Nehab's. There are two drawbacks to this approach. First, since normal maps are fields of unit vectors, the stitching must preserve unit length. Hence, the linear least squares solution used for textures would need to include quadratic equality constraints. This amounts to a quadratically constrained quadratic program which is no longer a convex optimisation problem. Second, the stitched texture will not necessarily correspond to a real surface. That is to say, the normals would not satisfy an integrability constraint.

We solve both of these problems by proposing a method to simultaneously stitch the normals and transfer the detail to the mesh. We do so using the same patch-based approach as for texture data and hence provide a unifying framework for Poisson blending both texture and shape using patches.

Our problem is closely linked to gradient-based [] or Laplacian-based [] mesh editing. However, our guidance field takes the form neither of a gradient field, nor of Laplacian (differential) coordinates. Nor does our guidance field arise through the editing of gradients or Laplacian coordinates of the original surface. Instead, we know the surface normal at each vertex where normals have been selected from different views. Also unlike mesh editing, we do not need to worry about local coordinate transformations. Our target normals are in world coordinates and we can update the mesh to directly match these normals.

We cannot directly apply mesh editing techniques to our problem. For the case of Laplacian coordinates, we would need to know the mean curvature normal at each vertex. Instead, we know only the unit surface normal. However, we can still obtain a linear system of equations by noting that the Laplacian coordinates and surface normal differ only by a scale factor:

$$\sum_{\{j|\{i,j\}\in\mathcal{K}\}} (\mathbf{v}_i - \mathbf{v}_j) \sim \mathbf{n}_i \tag{19}$$

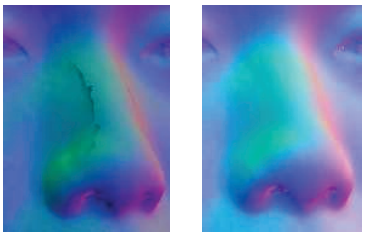
where $\sim$ denotes equality up to a non-zero scalar multiplication. Such sets of relations can be solved using the direct linear transformation (DLT) algorithm [?]. Accordingly, we write

$$[\mathbf{n}_i]_\times \sum_{\{j|\{i,j\}\in\mathcal{K}\}} (\mathbf{v}_i - \mathbf{v}_j) = \mathbf{0}, \tag{20}$$

where $\mathbf{0} = [0\ 0\ 0]^T$ and $[.]_\times$ is the cross product matrix:

$$[\mathbf{x}]_\times = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}. \tag{21}$$

We show in figure 7 an example of normals stitching results before and after applying the Poisson blending. Without the blending, the seams can be considerably visible on patches boundaries when the view they are selected from changes (e.g. on nose or cheek). Whereas



(a) (b)

Fig. 7. Normal stitching before (a) and after (b) Poisson blending

5.7 View Dependant Fresnel Gain

The measured specular albedo are subject to certain Fresnel-related effects depending on the angle between the surface normals and the view direction. These Fresnel gains increase with the view-angle and are particularly important at glancing angles where the view direction tends to be parallel to the surface.

In applications where the whole specular albedo is to be used, it is important to correct these Fresnel effect to achieve multi-view photometric consistency. Ghosh et al. [17] employed a data-driven method where they use the histogram of measured specular intensity in function of the view angle to scale each pixel intensity down to the average gain at zero-view angle. The authors assume that the specular reflectance parameters such as surface roughness are constant over the face and that the change in measured specular intensity is only conditioned by the Fresnel gain. Even though this method can yield less viewpoint-dependant specular albedo, the assumption of constant specular parameters reduces considerably the reliability of the result as an intrinsic quantity.

Our view merging method effectively prevents these issues. As stated above, the Fresnel gain is mostly noticeable at glancing angles where the deviation of the surface normal from the view vector approaches $\frac{\pi}{2}$. In our view merging process, the patch/view selection is such that the selected patches minimise, on average, the view angle. Thus the patches are generally picked from the view with minimal average angle, therefore, minimizing the chances to pick from regions at grazing angle.

6 Experimental Results

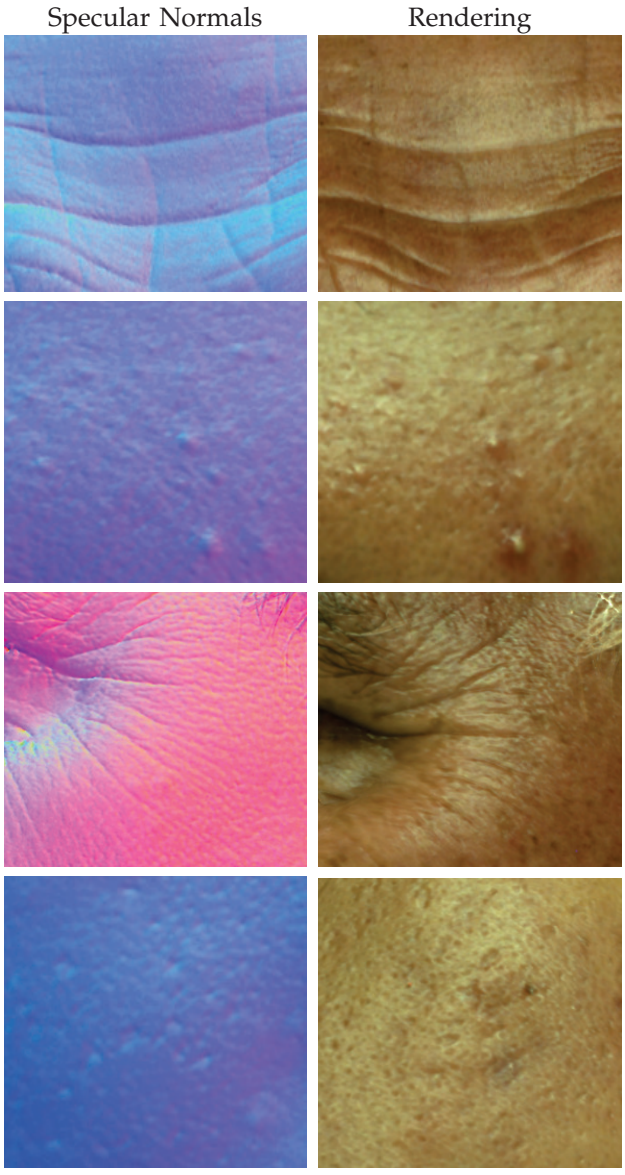


Fig. 8. Redering of Different Facial Regions

6.1 Views Merging

6.2 Detail Transfer

To refine the base mesh either the diffuse or specular surface normals can be used. Figure 10 shows the results of using each. One can notice that the diffuse normal maps tend to produce smother mesh while the specular normal maps yield more surface details. This is conform with the finding of Ma et al [1] and is explained by the fact that the surface reflectance from which the normals are estimated has different characteristics depending on whether it is diffuse or specular. In the diffuse case, the reflected light is considerably affected by subsurface scattering.

6.3 Rendering

We render using the Cook Torrance model [?] and the hybrid technique proposed by Ma et al. [1] where the estimated specular and diffuse surface normals are used to shade respectively the specular and diffuse components of the BRDF. We use two slopes of Beckmann functions to model the micro-facets distribution. In this work we assume constant roughness parameters and refraction index across the face. We give in figure 8 a few examples of rendering.

7 Conclusion

In this work we present a practical 3D face acquisition approach that allows capturing ear-to-ear mesh along with the skin micro-geometry and reflectance properties. Our system requires no prior geometrical calibration. The cameras parameters are obtained by structure-from-motion and are used to estimate a base mesh which is further refined using the recovered photometric surface normals. To achieve an ear-to-ear coverage of the face and overcome the problem of fixed photometric viewpoint inherent in polarized spherical gradient illumination, we capture the face in three poses and robustly stitch the corresponding normal maps and intrinsic textures into a seamless, complete and detailed face model.

Acknowledgments

We are grateful to Hadi Dahlan for assistance with data collection and to Fufu Fang for assistance with lightstage calibration and programming.

Reproducible Research

We make sample datasets available on the following webpage: ... These include the raw captured images, multiview and photometric stereo output as well as the final processed result.

References

[1] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, and P. Debevec, "Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination," in *Proc. Eurographics Symposium on Rendering*, 2007.

[2] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. SIGGRAPH*, 1999, pp. 187–194.

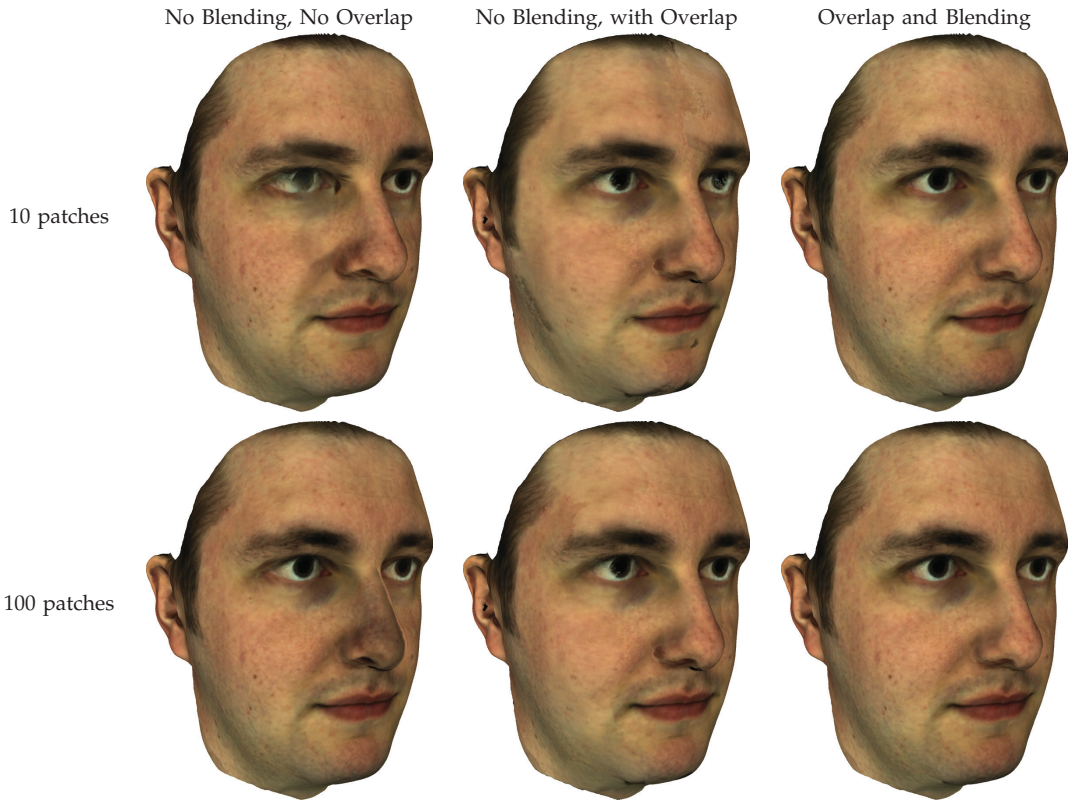


Fig. 9. Views Merging with Different Configurations

[3] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, 2003.

[4] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, 2003.

[5] F. Jiang, L. Dricot, V. Blanz, R. Goebel, and B. Rossion, "Neural correlates of shape and surface reflectance information in individual faces," *Neuroscience*, vol. 163, no. 4, pp. 1078–1091, 2009.

[6] J. D. Bustard and M. S. Nixon, "3D morphable model construction for robust ear and face recognition," in *Proc. CVPR*, 2010, pp. 2582–2589.

[7] A. Ghosh, T. Chen, P. Peers, C. A. Wilson, and P. Debevec, "Estimating specular roughness and anisotropy from second order spherical gradient illumination," *Computer Graphics Forum (Proceedings of EGSR)*, vol. 28, no. 4, pp. 1161–1170, 2009.

[8] ——, "Circularly polarized spherical illumination reflectometry," *ACM Trans. Graphic. (Proc. of SIGGRAPH Asia)*, vol. 29, no. 6, 2010.

[9] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith, and L. Smith, June 2011, pp. 132–139.

[10] D. Nehab, S. Rusinkiewicz, J. E. Davis, and R. Ramamoorthi, "Efficiently combining positions and normals for precise 3D geometry," *ACM Trans. Graphic. (Proceedings of SIGGRAPH)*, vol. 24, no. 3, pp. 536—543, 2005.

[11] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Opt. Eng.*, vol. 19, no. 1, pp. 139–144, 1980.

[12] C. A. Wilson, A. Ghosh, P. Peers, J.-Y. Chiang, J. Busch, and P. Debevec, "Temporal upsampling of performance geometry using photometric alignment," *ACM Trans. Graphic. (Proceedings of SIGGRAPH)*, vol. 29, no. 2, 2010.

[13] E. Claridge, S. Cotton, P. Hall, and M. Moncrieff, "From colour to tissue histology: physics based interpretation of images of pigmented skin lesions," *Med. Image Anal. J.*, vol. 7, pp. 489–502, 2003.

[14] Y. Furukawa and J. Ponce, "Dense 3d motion capture from synchronized video streams," in *Proc. CVPR*, 2008.

[15] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross, "High-quality single-shot capture of facial geometry," *ACM Trans. Graphic. (Proceedings of SIGGRAPH)*, vol. 29, no. 3, 2010.

[16] C. Wu, Y. Liu, Q. Dai, and B. Wilburn, "Fusing multiview and photometric stereo for 3D reconstruction under uncalibrated illumination," *IEEE Trans. Vis. Comp. Gr.*, vol. 17, no. 8, pp. 1082–1095, 2011.

[17] A. Ghosh, G. Fyffe, B.Tunwattanapong, J. Busch, X. Yu, and P. Debevec, "Multiview face capture using polarized spherical gradient illumination," *ACM Trans. Graphic.*, vol. 30, no. 6, p. 129, 2011.

[18] J. Park, S. n Sinha, Y. Matsushita, Y.-W. Tai, and I. S. Kweon, "Multiview Photometric Stereo using Planar Mesh Parameterization," in *ICCV*. International Conference on Computer Vision, December 2013. [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=207997

[19] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *J. Opt. Soc. Am. A*, vol. 11, no. 11, pp. 3079–3089, Nov 1994. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-11-11-3079

[20] S. Mallick, T. Zickler, D. Kriegman, and P. Belhumeur, "Beyond Lambert: reconstructing specular surfaces using color," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, June 2005, pp. 619–626 vol. 2.

[21] M. Chuang, L. Luo, B. J. Brown, S. Rusinkiewicz, and M. Kazhdan, "Estimating the Laplace-Beltrami operator by restricting 3D functions," *Comput. Graph. Forum*, vol. 28, no. 5, pp. 1475–1484, Jul. 2009.

[22] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graphic. (Proceedings of SIGGRAPH)*, vol. 22, no. 3, pp. 313–318, 2003.

[23] G. Peyré and L. D. Cohen, "Geodesic remeshing using front propagation," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 145–156, Aug. 2006.

[24] J. Totz, A. J. Chung, and G.-Z. Yang, "Patient-specific texture blending on surfaces of arbitrary topology," in *Workshop on Augmented environments for Medical Imaging and Computer-aided Surgery*, London, UK, Sep. 2009, pp. 78–85.

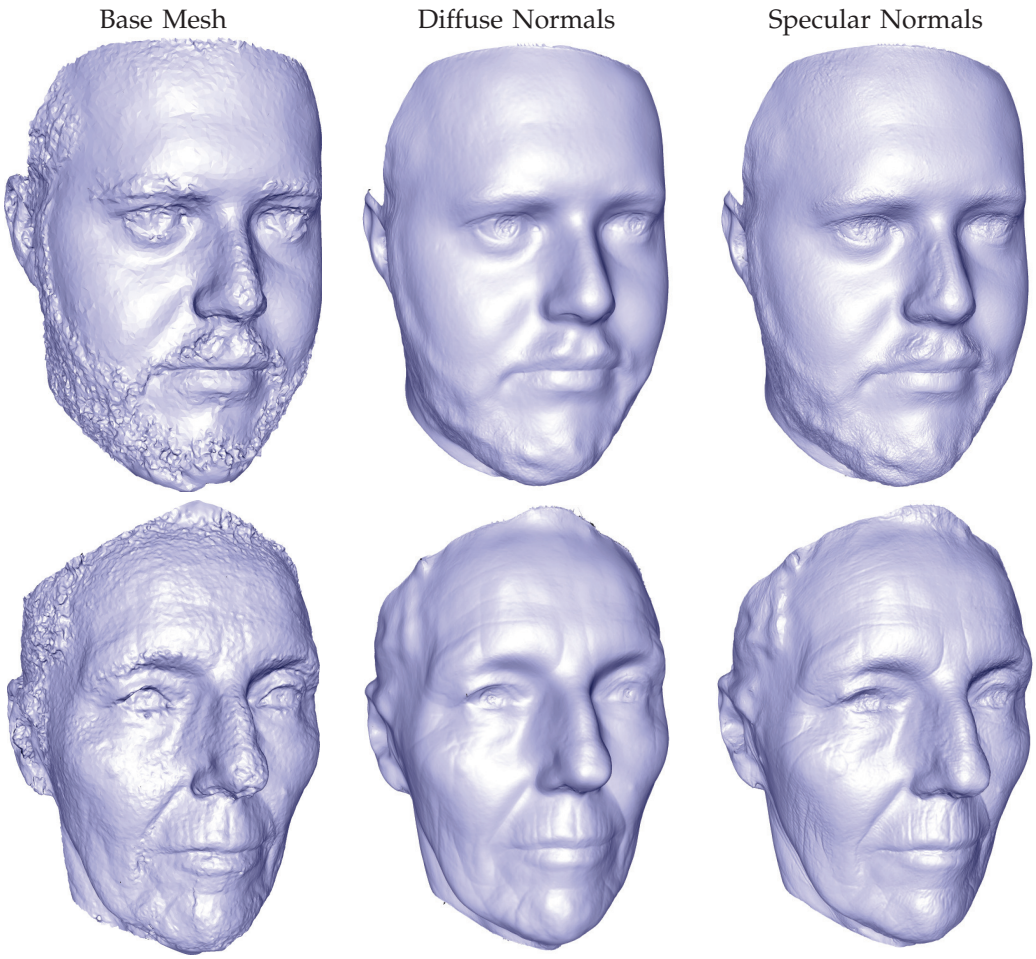


Fig. 10. Mesh Refinement Using Diffuse Normals (middle column) or Specular Normals (right column)

Alassane Seck received an M.sc. in Artificial Intelligence (Language and Image Processing) from the University of Caen, France and an M.sc in Image Processing and Remote Sensing from ENSEGID, Bordeaux, France. He is currently a PhD student in Computer Vision at Aberystwyth University. His research interests include Computer Vision, Computer Graphics, 3D surface texture modelling, 3D surface micro-structure measurements and Machine Learning.

Abhishek Dutta received the Bachelor's degree in Computer Engineering from Tribhuvan University (Nepal) in 2009 and the Master of Science (by research) in Computer Science from the University of York (UK) in 2012. He is currently working toward the PhD degree in Computer Science at the University of Twente (Netherlands). His research interests are in the area of Computer Vision, Computer Graphics and Machine Learning.

William A. P. Smith (M'08) received the B.Sc. degree in computer science, and the Ph.D. degree in computer vision from the University of York, York, U.K. He is currently a Lecturer with the Department of Computer Science, University of York, York, U.K. His research interests are in face modeling, shape-from-shading, reflectance analysis and the psychophysics of shape-from-X. He has published more than 70 papers in international conferences and journals, was awarded the Siemens best security paper prize at BMVC 2007, and was finalist as the U.K. nominee for the ERCIM Cor Baayen award 2009. He is an associate editor of the IET journal Computer Vision, and has served as co-chair of the ACM International Symposium on Facial Analysis and Animation in 2010 and 2012, and the CVPR 2008 workshop on 3D Face Processing. He is a member of the IEEE and BMVA.

Bernie Tiddeman is a Senior Lecturer and Head of Department in Computer Science at Aberystwyth University. He obtained his BSc from University of St Andrews in 1992, MSc from Manchester University in 1994 and PhD from Heriot-Watt University in 1998. From 1999-2010 he worked as a researcher and then lecturer at the University of St Andrews. His research interests include 2D and 3D facial image analysis and synthesis, including texture modelling for age estimation and age progression and skin health analysis and synthesis.

Hannah M. Dee is a lecturer in computer science at Aberystwyth University in the UK. Before this, she carried out post-doctoral work at the Grenoble Institute of Technology (INPG), the University of Leeds, and Kingston University. She received her Ph.D. in computer vision from the University of Leeds in 2006. Her research area is computer vision, with a particular interest in vision for modelling change, growth and texture. She is deputy chair of BCSWomen, the British Computer Societys group for women in technology, and is active in encouraging women and girls to consider careers in computer science.

Arnaud Dessein received the Dipl.-Ing. degree from École Centrale de Lille, Lille, France, the M.Sc. degree in acoustics, signal processing and computer science applied to music, and the Ph.D. degree in computer science from Université Pierre et Marie Curie, Paris, France. He is currently a Research Associate with the Department of Computer Science, University of York, York, U.K. His research interests include computer vision, audio analysis, signal processing, machine learning, statistics and information theory. He has authored 1 book chapter, 2 international journal articles and 5 refereed conference papers. He is a member of SEE.

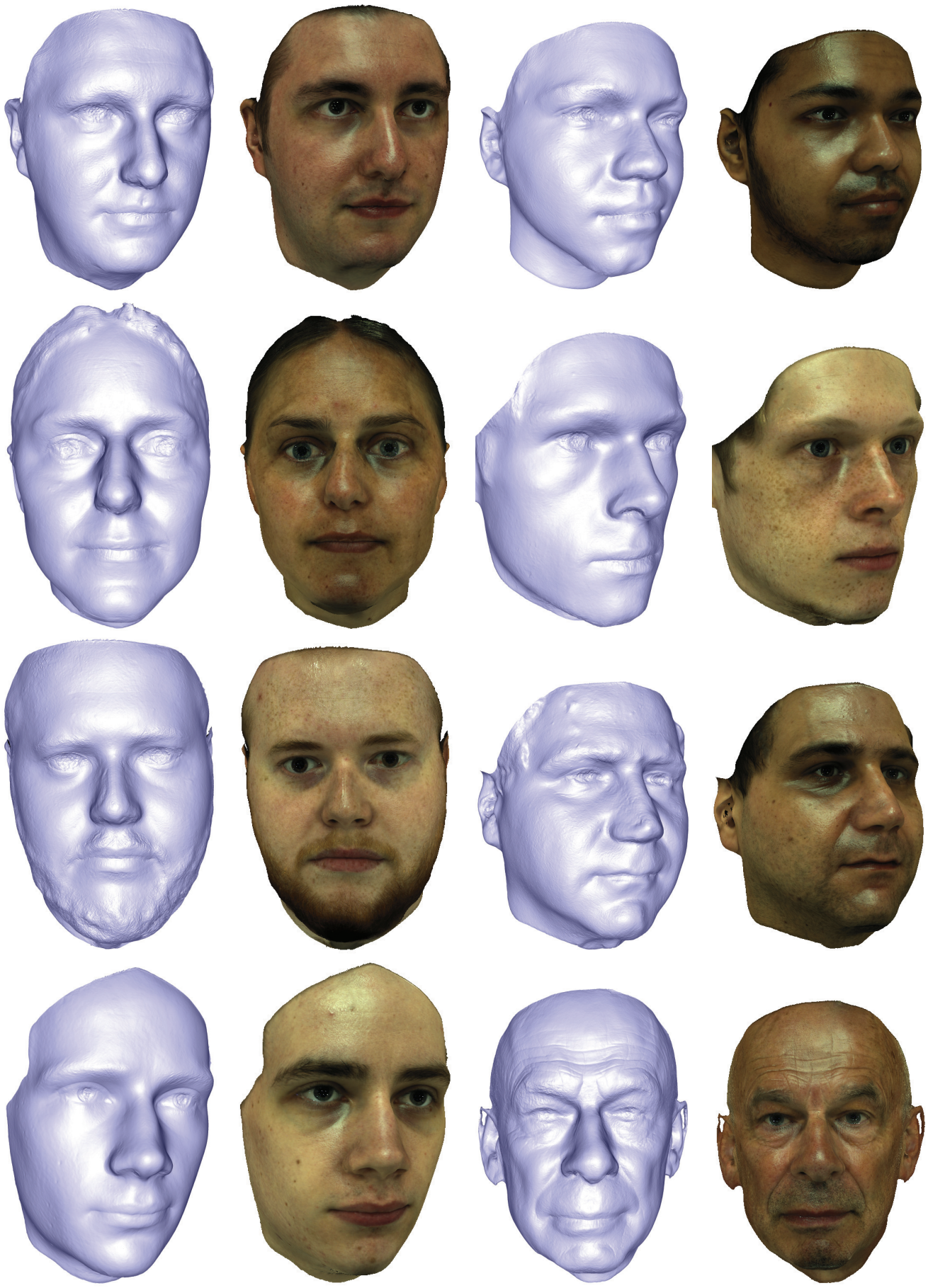


Fig. 11. Geometry mesh and Renderings of Different subjects